Abstract

The web is a huge repository of knowledge and numerous hyperlinks. Web also serves a broad diversity of user communities and global information service centers. Every day the knowledge in web page upwards rapidly. Web pages can be used to convey the knowledge to web users. Such voluminous size of the web makes an intricacy of web information retrieval, web content filtering and web structure mining. Hence, it is essential for proper categorization of web pages. This paper demonstrates the web page categorization problem as the multi classification task and provides a suitable solution using a supervised learning technique namely multilayer perceptron. The classification model is generated by learning the features that have been extracted from HTML structure and URL of the web page. Feature reduction techniques have been applied to select optimum features and a model is learned. The experimental results of the multilayer perceptron models before and after feature reduction has been evaluated and observed that the multilayer perceptron model with reduced features performs well.

References

Web Page Categorization using Multilayer Perceptron with Reduced Features

- Arabib and Michael A, The Handbook of Brain Theory and Neural Networks.
- Shiqun Yin, Yuhui Qiu, Chengwen, Zhong, Jifu Zhou. 2007. Study of Web Information
Web Page Categorization using Multilayer Perceptron with Reduced Features


Index Terms

Computer Science                Information Sciences

Keywords
Categorization  Multilayer perceptron  Training  Web page